EOSC
European Open Science Cloud

What is it? What can it do for us?

Fabio Pasian – INAF-OATs

ICT Workshop, Trieste, 17 November 2016

European Cloud Initiative to give Europe a global lead in the data-driven economy

Brussels, 19 April 2016

The Commission today presented its blueprint for cloud-based services and world-class data infrastructure to ensure science, business and public services reap benefits of big data revolution.

Europe is the largest producer of scientific data in the world, but insufficient and fragmented infrastructure means this 'big data' is not being exploited to its full potential. By bolstering and interconnecting existing research infrastructure, the Commission plans to create a new **European Open Science Cloud** that will offer Europe's 1.7 million researchers and 70 million science and technology professionals a virtual environment to store, share and re-use their data across disciplines and borders. This will be underpinned by the **European Data Infrastructure**, deploying the high-bandwidth networks, large scale storage facilities and super-computer capacity necessary to effectively access and process large datasets stored in the cloud. This world-class infrastructure will ensure Europe participates in the global race for high performance computing in line with its economic and knowledge potential.

Focusing initially on the scientific community - in Europe and among its global partners -, the user base will over time be enlarged to the public sector and to industry. This initiative is part of a package of measures to strengthen Europe's position in data-driven innovation, to improve competitiveness and cohesion and to help create a Digital Single Market in Europe (press release).

Carlos **Moedas**, Commissioner for Research, Science and Innovation, said: "*Our goal is to create a*

By bolstering and interconnecting existing research infrastructure, the Commission plans to create a new **European Open Science Cloud** that will offer Europe's 1.7 million researchers and 70 million science and technology professionals a virtual environment to store, share and re-use their data across disciplines and borders. This will be underpinned by the **European Data Infrastructure**, deploying the high-bandwidth networks, large scale storage facilities and super-computer capacity necessary to effectively access and process large datasets stored in the cloud. This world-class infrastructure will ensure Europe participates in the global race for high performance computing in line with its economic and knowledge potential.

Investment (!!!)

The public and private investment needed to implement the European Cloud Initiative is estimated at **€6.7 billion**. The Commission estimates that, overall, **€2 billion** in Horizon 2020 funding will be allocated to the European Cloud initiative. The estimation of the required additional public and private investment is **€4.7 billion** in the period of 5 years.

In addition to the European research community, the European Open Science Cloud and the European Data Infrastructure will be accessible and bring benefits for a host of other users:

- **Businesses** will have cost-effective and easy access to top level data and computing infrastructure, as well as a wealth of scientific data enabling data-driven innovation. This will particularly benefit **SMEs**, which typically lack access to such resources.
- **Industry** will benefit from the creation of a large-scale cloud eco-system, supporting the development of new European technologies such as low-power chips for high performance computing.
- **Public services** will benefit from reliable access to powerful computing resources and the creation of a platform to open their data and services, which can lead to cheaper, better and faster interconnected public services. Researchers will also benefit from online access to the wealth of data created by public services.

The Commission will progressively put in place the European Cloud Initiative through a series of actions, including:

- As of 2016: creating a **European Open Science Cloud** for European researchers and their global scientific collaborators by integrating and consolidating e-infrastructure platforms, federating existing scientific clouds and research infrastructures, and supporting the development of cloud-based services.

- 2017: **opening up by default all scientific data** produced by future projects under the €77 billion Horizon 2020 research and innovation programme, to ensure that the scientific community can re-use the enormous amount of data they generate.

- 2018: launching a flagship-type initiative to accelerate the nascent development of **quantum technology**, which is the basis for the next generation of supercomputers.

- By 2020: developing and deploying a large scale **European high performance computing, data storage and network infrastructure**, including by acquiring two prototype next-generation supercomputers of which one would rank among the top three in the world, establishing a European big data centre, and upgrading the backbone network for research and innovation (GEANT).

The Commission will progressively put in place the European Cloud Initiative through a series of actions, including:

- As of 2016: creating a **European Open Science Cloud** for European researchers and their global scientific collaborators by integrating and consolidating e-infrastructure platforms, federating existing scientific clouds and research infrastructures, and supporting the development of cloud-based services.

- 2017: **opening up by default all scientific data** produced by future projects under the €77 billion Horizon 2020 research and innovation programme, to ensure that the scientific community can re-use the enormous amount of data they generate.

- 2018: launching a flagship-type initiative to accelerate the nascent development of **quantum technology**, which is the basis for the next generation of supercomputers.

- By 2020: developing and deploying a large scale **European high performance computing, data storage and network infrastructure**, including by acquiring two prototype next-generation supercomputers of which one would rank among the top three in the world, establishing a European big data centre, and upgrading the backbone network for research and innovation (GEANT).

Bootstrapping EOSC

- *"As of 2016: creating a **European Open Science Cloud** for European researchers and their global scientific collaborators …"*

- October 2016: Recommendations of the Commission <u>High Level Expert Group</u> on the European Open Science Cloud

KDD Lab, UniPI

  - *Members : Paul Ayris, Jean-Yves Berthou, Rachel Bruce (Rapporteur), Stefanie Lindstaedt, Anna Monreale, Barend Mons (Chair), Yasuhiro Murayama (Observer, Japan), Caj Södergård, Klaus Tochtermann, Ross Wilkinson (Observer, Australia).*

  - draft available in April

DATA SHARING

LINKING DATA

CONNECTING SCIENTISTS

SUSTAINABLE

OPEN SERVICES

CONNECTING DISCIPLINES

BETTER SCIENCE

Research and Innovation

EOSC – HLEG recommendations

- Close discussions about the 'perceived need' of a science cloud and to take immediate action on the EOSC in close concert with Member States building on existing capacity and expertise.

- Writing clear Rules of Engagement for access to the EOSC and for the provision of services based on research data

- Framing the EOSC as the EU contribution to a future, global Internet of FAIR Data and Services underpinned by open protocols.

- Set-up and fund a concerted effort to develop core data expertise in Europe. Estimation: half a million 'core data scientists' are needed to make the most of open research data in Europe.

- Change radically the funding model for research data, from traditional and rigid funding schemes of the past - e.g. small and unaccounted part of a time-limited and space-bound grants to an overall co-funded national / EC funding scheme. Estimation: on average about 5% of total research expenditure should be spent on properly managing and 'stewarding' data in an integrated fashion.
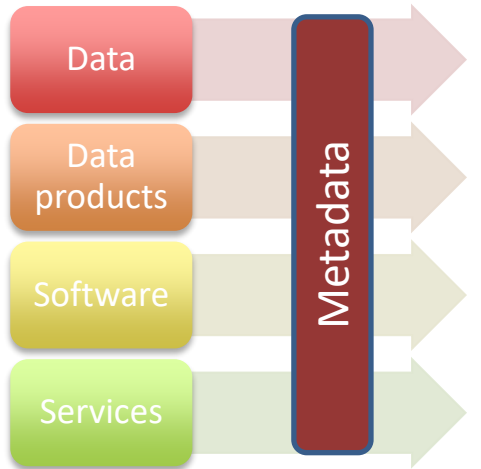
Is it

European <u>Open Science</u> Cloud

or

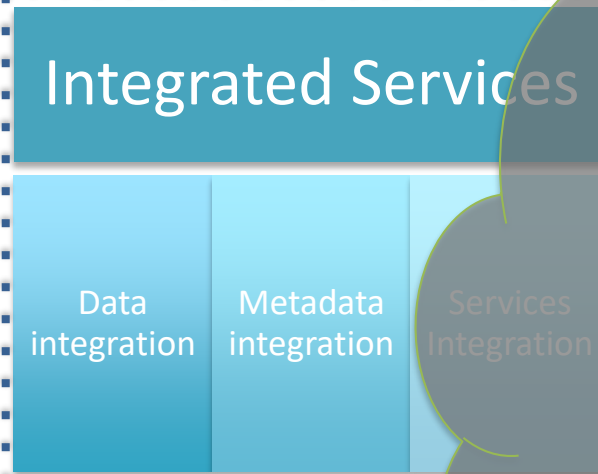European Open <u>Science Cloud</u>

?

Reasonably both:

1. European Open <u>Science Cloud</u>

2. European <u>Open Science</u> Cloud
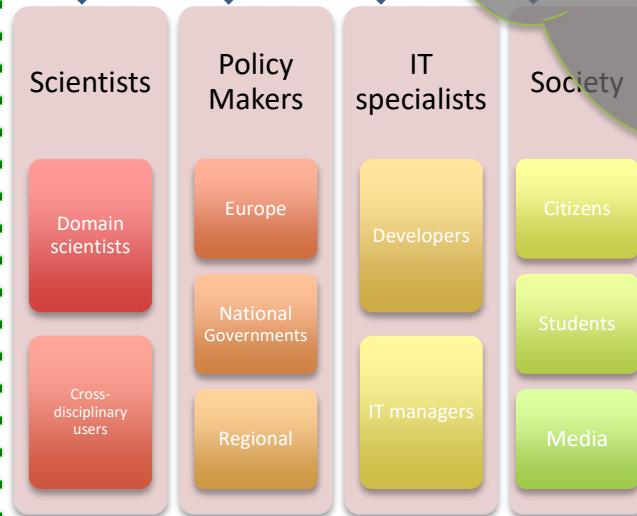
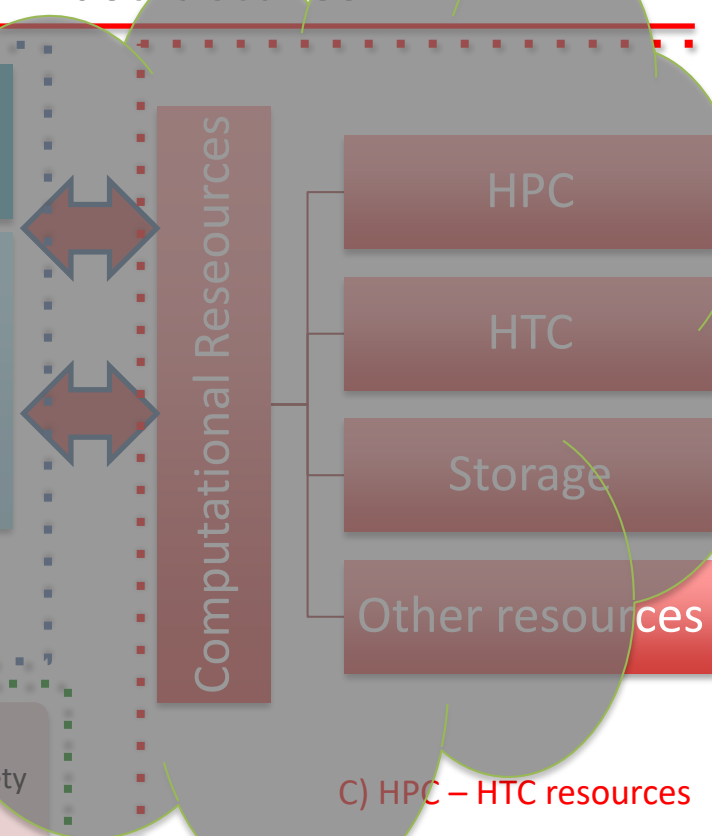**Research Infrastructures** — **e-Infrastructures**

Data
Data products
Software
Services

Metadata

A) data generation

Integrated Services

Data integration | Metadata integration | Services Integration

B) data & service integration

Computational Reseources

HPC
HTC
Storage
Other resources

C) HPC – HTC resources

Data Lifecycle

Production
Analysis
Publication
Archive
Re-use

Scientists | Policy Makers | IT specialists | Society

Domain scientists | Europe | Developers | Citizens
Cross-disciplinary users | National Governments | IT managers | Students
 | Regional | | Media

D) engagement & communication
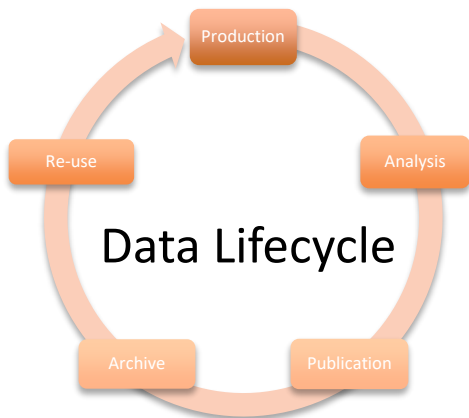
Users and Stakeholders

What will EOSC cover ?

If EOSC extends towards integrated services, it requires an increased RI influence on governance, technology, use cases etc.

12

EOSCpilot: the proposal

- *"As of 2016: creating a **European Open Science Cloud** for European researchers and their global scientific collaborators …"*
- Proposal
  - in response to INFRADEV-04-2016 call
  - RIA (Research and Innovation Action)
  - submitted 22/06/2016 16:47:52
  - multi-domain
  - 33 partners, led by UK STFC
  - duration 24 months
  - budget ~ 10M€
  - reached negotiation phase

EOSCpilot: objectives (I)

From the Abstract of the proposal:

The EOSCpilot project will support the first phase in the development of the European Open Science Cloud (EOSC) as described in the EC Communication on European Cloud Initiatives [2016].

- It will establish the governance framework for the EOSC and contribute to the development of European open science policy and best practice;

- It will develop a number of pilots that integrate services and infrastructures to demonstrate interoperability in a number of scientific domains; and

- It will engage with a broad range of stakeholders, crossing borders and communities, to build the trust and skills required for adoption of an open approach to scientific research.

…

EOSCpilot: objectives (II)

... These actions will build on and leverage already available resources and capabilities from research infrastructure and e-infrastructure organisations to maximise their use across the research community. The EOSCpilot project will address some of the key reasons why European research is not yet fully tapping into the potential of data. In particular, it will:

- reduce fragmentation between data infrastructures by working across scientific and economic domains, countries and governance models, and

- improve interoperability between data infrastructures by demonstrating how data and resources can be shared even when they are large and complex and in varied formats.

In this way, the EOSCpilot project will improve the ability to reuse data resources and provide an important step towards building a dependable open-data research environment where data from publicly funded research is always open and there are clear incentives and rewards for the sharing of data and resources.

EOSCpilot: structure
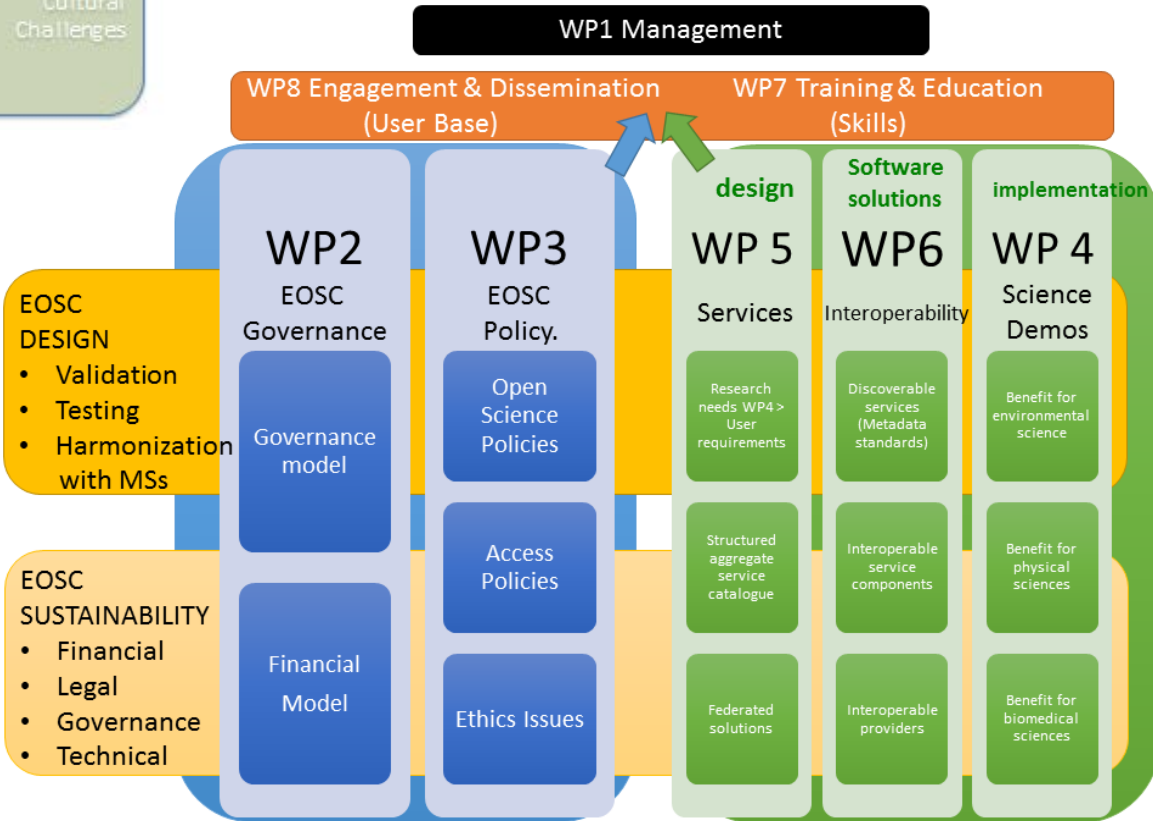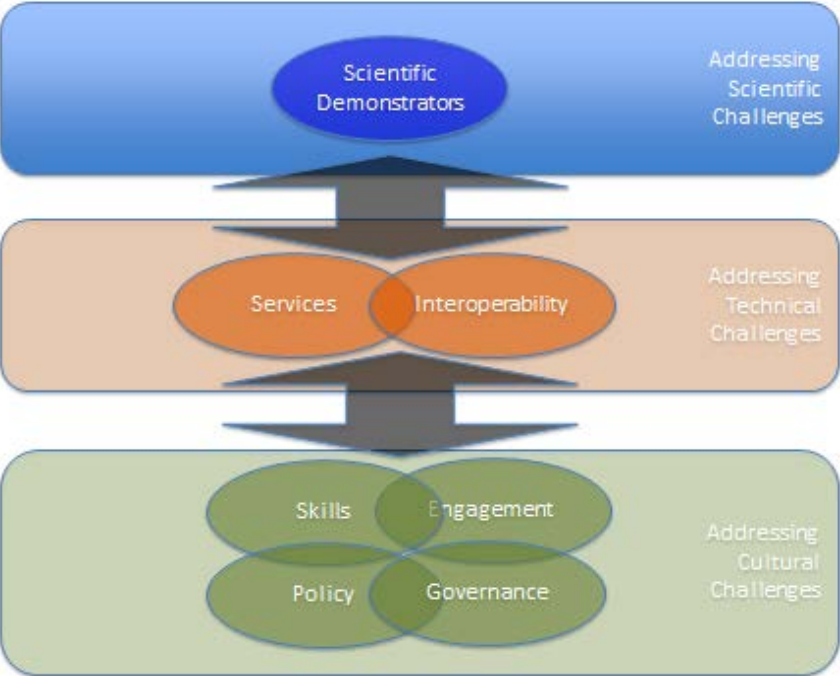
Astro in EOSCpilot

- … almost missing! ☹
- Proposal was generated by grouping of disciplinary clusters → ASTERICS was involved
  - mandate from ESFRIs: monitor developments to evaluate potential impact
  - ASTERICS PM and AGA Chair participated
  - no astro science demos approved in proposal
- Small participation (INAF ☺) in WP5 (Services) and WP6 (Interoperability)

INAF Background

- HTC services for science since 2001
  - Grid.IT, DRACO, IGI, EGEE projects (I-II-III), EGI projects (Inspire, Engage)
- Distribute computing resources for Astronomy (Linux based, common services e.g. AAI, storage, monitoring)
- Coordination of European A&A Community in EGEE and EGI (since 2008)

INAF-OATs distributed computing Lab

- HTC cluster based on EMI grid middleware:
  - 100 Cores, 16 TB storage, monitoring and accounting
  - Support: High Energy Astrophysics, Radio Astronomy
  - Closed in April 2016
- Cloud Resource:
  - 200 Cores, RAM 8GB/cores, 70TB storage, Openstack
  - Open to INAF for testing and pre-production
  - Integrated with INAF AAI (CED)
  - Integrated in EGI.eu federated cloud.
- hotCat (DHTCS-IT: ~1000 Cores, …) → CHIPP

INAF in EOSCpilot (I)

- WP5 (Services):
  - Definition of a portfolio of validated EOSC services and the related rules of engagement for service providers in order to define the activities and processes regulating their provisioning and allow a further future enhancement and evolution of the portfolio.
  - INAF (6 PM) gives input to
    1. definition of the overall EOSC architecture
    2. definition of the EOSC service portfolio
    3. definition of the EOSC federated service management

INAF in EOSCpilot (II)

- ## WP6 (Interoperability):
  - The interoperability WP develops and demonstrates the interoperability requirements between e-Infrastructures, domain research infrastructures and other service providers needed in the European Open Science Cloud.
  - INAF (6 PM) gives input to
    - Sub-task 6.2.3 Data Interoperability (as part of the FAIR principle): co-responsible of preparation of final report

- ## You are welcome to provide your input !

Things to take home

- EOSC is going to happen!
  - the EC is investing a lot in this initiative
  - there will be no other significant EC funding for clouds
- EOSC will be complementary to the dedicated facilities for the big projects → "long tail of science"
- EOSC will **NOT** be Astro-driven!
  - but, as a community, we have a chance of providing input
- There is no free lunch!
  - either we stay out
  - or we pay for the usage of EOSC
  - or we contribute to EOSC with appropriate computing & storage (& knowledge!) resources
- Needed: education + careers for data scientists

Thank you !
Any question ?

fabio.pasian@inaf.it